GoogleBooks discussion paper for CONSER Operations Committee meeting (2009)
Steve Shadle & Cindy Hepfer with contributions by Robert Bremer

Background

As part of OCLC's efforts to synchronize WorldCat with important mass digitization projects such as Google Book Search (GBS) and MS Live Search Books, bibliographic records representing the online version of books and serials have been loaded into WorldCat. The initial phase of the Google Book (GB) effort involved the creation of more than 800,000 records for monographs and more than 40,000 records for serials. Staff working on the project applied the eMonograph model for cloning print records resulting in provider-specific records with reproduction notes that were added to WorldCat without matching. While adequate for eMonographs, this approach was incorrect for eSerials which are cataloged according to a provider-neutral policy and require matching against existing WorldCat records.

With the pending adoption of the provider-neutral cataloging approach for eMonographs, processing of GB titles will need to change to be more in line with eSerials. Existing eSerial records for GB titles will be made provider-neutral and will be allowed to match other eSerials record in WorldCat in the pending global duplicate detection project.

Discussion

Depending on the participating library's local record which is used as the basis for the GB record, the resulting GB record may or may not follow current cataloging practice. Short term, CONSER should develop a common understanding of how to treat GB eSerial records within the context of existing CONSER practice and should make any recommendations for changes in GB record processing. Longer term, CONSER should examine the role of preservation and digitization project data in the CONSER record.

Process (How an OCLCE/Google Books Record Gets Into WorldCat)

- Participating library provides Google with local print version records
- Google adds/edits a set of fields to these records and submits to OCLC
- Fields added/edited include:
    - ELvl: K
    - Form: s
    - 040 OCLCE $b eng OCLCE
    - 050 indicator values
    - 245 $h [electronic resource]
    - 506, 530, 533 added
    - Reproduction 776 added pointing to local library record
    - 856: one 856 added for each physical volume digitized with the following subfields: $3 volume marking $u Google URL $z generic GB note

Example Records (University of Michigan print version and resulting GB record)

LDR	nas 22002171 4500
005	19880719000000.0
006	m d
007	cr bn ---auaua
008	880719d19591962caumu p 0uuu 0eng d
035	|a (RLIN)MIUG07456-S
035	|a (CaOTULAS)160562000
040	|a CU-CU |c CU-CU |d MiU
24500	|a CA; |b the journal of commercial art.
260	|a Palo Alto, Calif.
300	|b ill.
3620	|a v. 1-4, no. 3; 1959-1962.
500	|a Subtitle varies.
538	|a Mode of access: Internet.
650 0	|a Commercial art |x Periodicals.
78500	|t CA magazine

```
Type a    ELvl K    Srce d    GPub      Ctrl      Lang eng
BLvl s    Form s    Conf 0    Freq m    MRec      Ctry cau
S/L  0    Orig      EntW      Regl u    Alph
Desc      SrTp p    Cont      DtSt d    Dates 1959,1962
```

007 c ǂb r ǂd u ǂe n ǂf u ǂg --- ǂh u ǂi u ǂj a ǂk u ǂl a

040 OCLCE ǂb eng ǂc OCLCE

245 00 CA; **ǂh [electronic resource]** ǂb the journal of commercial art.
260 Palo Alto, Calif.

300 ǂb ill.
362 0 v. 1-4, no. 3; 1959-1962.
500 Subtitle varies.

506 Access to some volumes or items may be restricted ǂf Restrictions unspecified ǂ2 star
530 Also issued in print.

533 Electronic reproduction. ǂb Mountain View, Calif. : ǂc Google Book Search Library Project, ǂd 2009. ǂn Mode of access: World Wide Web.

650 0 Commercial art ǂx Periodicals.

785 00 ǂt CA magazine

776 1 ǂc Original ǂw (OCoLC)243894306

856 40 ǂ3 v.3 1961 Jul-Dec ǂu http://books.google.com/books?id=QNtUAAAAMAAJ ǂz Google Book Search Library Project

856 40 ǂ3 v.3 1961 Jan-Jun ǂu http://books.google.com/books?id=mdpUAAAAMAAJ ǂz Google Book Search Library Project

856 40 ǂ3 v.4 1962 Jul-Dec ǂu http://books.google.com/books?id=_NtUAAAAMAAJ ǂz Google Book Search Library Project

856 40 ǂ3 v.2 1960 Jan-Jun ǂu http://books.google.com/books?id=4dlUAAAAMAAJ ǂz Google Book Search Library Project

856 40 ǂ3 v.1 ǂu http://books.google.com/books?id=_9hUAAAAMAAJ ǂz Google Book Search Library Project

856 40 ǂ3 v.2 1960 Jul-Dec ǂu http://books.google.com/books?id=NNpUAAAAMAAJ ǂz Google Book Search Library Project

Note: Because there was no corresponding OCLC record identified in the original print version record, that record was also loaded as an Encoding Level 'M' record (#243894306). This record number is included in the 776 link pointing from the eSerial record to the print version record. Data in bold face was added as part of GB-load process.

Questions/Short Term

These questions arose during CONSRLST discussions beginning February 2009. Most have been tentatively resolved and only need confirmation, some require further discussion. If necessary, this list of questions can serve as a basis for a CONSER FAQ on the topic.

Can CONSER members edit GB records for eSerials to reflect aggregator-neutral practice? Can we authenticate them?

- Robert Bremer (RB): Google/OCLCE records are intended to be aggregator-neutral serial records like any other eSerial records and are not a category of allowed duplicates. CONSER libraries can edit and authenticate them or report them as duplicates of other eSerial records.

Can CONSER members delete the multiple 856s that correspond to the print volumes of the participating library? Are there any alternatives for handling the GB URLs?

- Some libraries have reported up to three pages of GB 856s. Discussion list consensus is that we are stuck with keeping the links for now. Some periodicals that have been digitized as "magazines" (through Google's publisher partner program) have a separate URL for each issue and/or selected articles (rather than bound volume) but do provide browsing the entire run from every issue/article page. These URLs have not been submitted to WorldCat. Examples:
    - Ebony, July 2004: http://books.google.com/books?id=JdkDAAAAMBAJ
    - Jet, 4/30/84: http://books.google.com/books?id=D7MDAAAAMBAJ

When a serial is available through the publisher partner program, can a single issue URL be used in lieu of multiple URLs corresponding to the participating library's bound volumes? Is it necessary to keep all GB URLs if participating libraries overlap in their digitization?

Should GB records for eSerials be reported as duplicates?

- RB: Google/OCLCE records are not a category of allowed duplicates and should be reported as such. The only issue is misplaced Google URLs, since deleting them outright would be a big problem. For any given eSerial, CONSER libraries may move the Google URLs to the corresponding record or they can report the "problem" to OCLC Quality Control staff to handle.

Is GBS considered a preservation project and thus tagging practices should be similar to the *Registry of Digital Masters*? Should the 506 and 533 fields be retained?

- RB: GBS is not considered a preservation project/service so use of 506/533 is not appropriate.

There are some other differences in tagging between GB records and CONSER aggregator-neutral. These include lack of ER 006, use of 530/776 rather than 776 08, and use of reproduction 776. Should CONSER catalogers accept these as is or edit the GBS record?

- RB: Most of these were due to inappropriate application of the eBook macro to serial records. OCLC staff will clean up existing GB eSerial records to conform to current CONSER practice and the processing of future loads will be adjusted so records will conform to CONSER aggregator-neutral practice. CONSER catalogers can currently edit/delete these fields as they would any other eSerial record they are authenticating.

Renette Davis: *Can these edits include edits to 856$3 to clarify part linked to?*

The GB record is pre-AACR2/latest-entry/minimal/etc. I would like to authenticate the eSerial record, but unsure whether I'm allowed to redescribe (either on what's available from Google or from print version record)...

- Since OCLC has stated these records should be treated like any other eSerial records, catalogers should make the same judgments they would when working with any other OCLC copy. The significant difference is that all appropriate GB 856 fields must be retained.

What happens if GBS digitizes more issues of a serial? How does record maintenance work?

- RB: There is a behind-the-scenes table that tracks which individual print version records resulted in the addition of particular online version records. If additional volumes of a monograph or issues of serial are digitized at a different point in time, the table is used to determine whether a print record has already been cloned and if so the additional information is added to the online version record already in the database.

Questions/Long-Term

GB has raised questions related to the use of OCLC/CONSER records as a vehicle to carry digitization/preservation project-specific information. Existing projects discussed to date are the *Registry of Digital Masters* and *Print Preservation Project*).

- Is the national bibliographic record the best place to house this data?
- In practical terms, are there are rules of thumb CONSER can agree on for deciding which fields to include on the national bibliographic record or for considering what is a 'digitization' project vs. a 'preservation' project?
- Can we envision a different place for recording and maintaining digital preservation project data?
    - Local Holdings Records?
    - Institutional Records (most of Stanford's loaded records are IRs)?
    - Web-based pages (similar to OpenLibrary for books openlibrary.org or WorldCat Identities for authors www.worldcat.org/identities/)?

Finally, in considering the OCLC eContent Synchronization program, other mass digitization efforts, the potential for inconsistency between print and online record 856s and resulting duplicate 856 maintenance, is it time to revisit CONSER's single-record approach?